Denoising Raw Photon Counting Atmospheric Lidar Data Using Autoencoders

Patrick Selmer[1], John Yorks[2], and Edward Nowottnick[2]

[1] Science Systems and Applications, Inc.
[2] NASA GSFC

Background and Motivation

Atmospheric lidars provide critical information about the vertical distribution of clouds and aerosols thereby improving our understanding of the climate system. Many lidars use photon counting detectors. These detectors are highly sensitive, allowing for the detection of optically thin clouds and aerosols. However, during the day the solar background signal can be much greater than the signal from these features, making them difficult to detect. Averaging the data horizontally across profiles has been the standard way to increase SNR, but at the expense of resolution. Modern, Deep Learning based denoising algorithms can be applied to improve the SNR without coarsening resolution.

Method

1. Identify pairs of noisy and clean image samples
   - Use night data, add artificial noise
   - 3 noise levels chosen following experiments outlined in *Zhang et al 2017* and *Jia et al 2021*
2. Choose an autoencoder-style Neural Network architecture
3. Train model to predict residual ***noise*** instead of clean image. DL models are easier to train this way [*Zhang et al 2017*].

Two image denoising model architectures were tried:

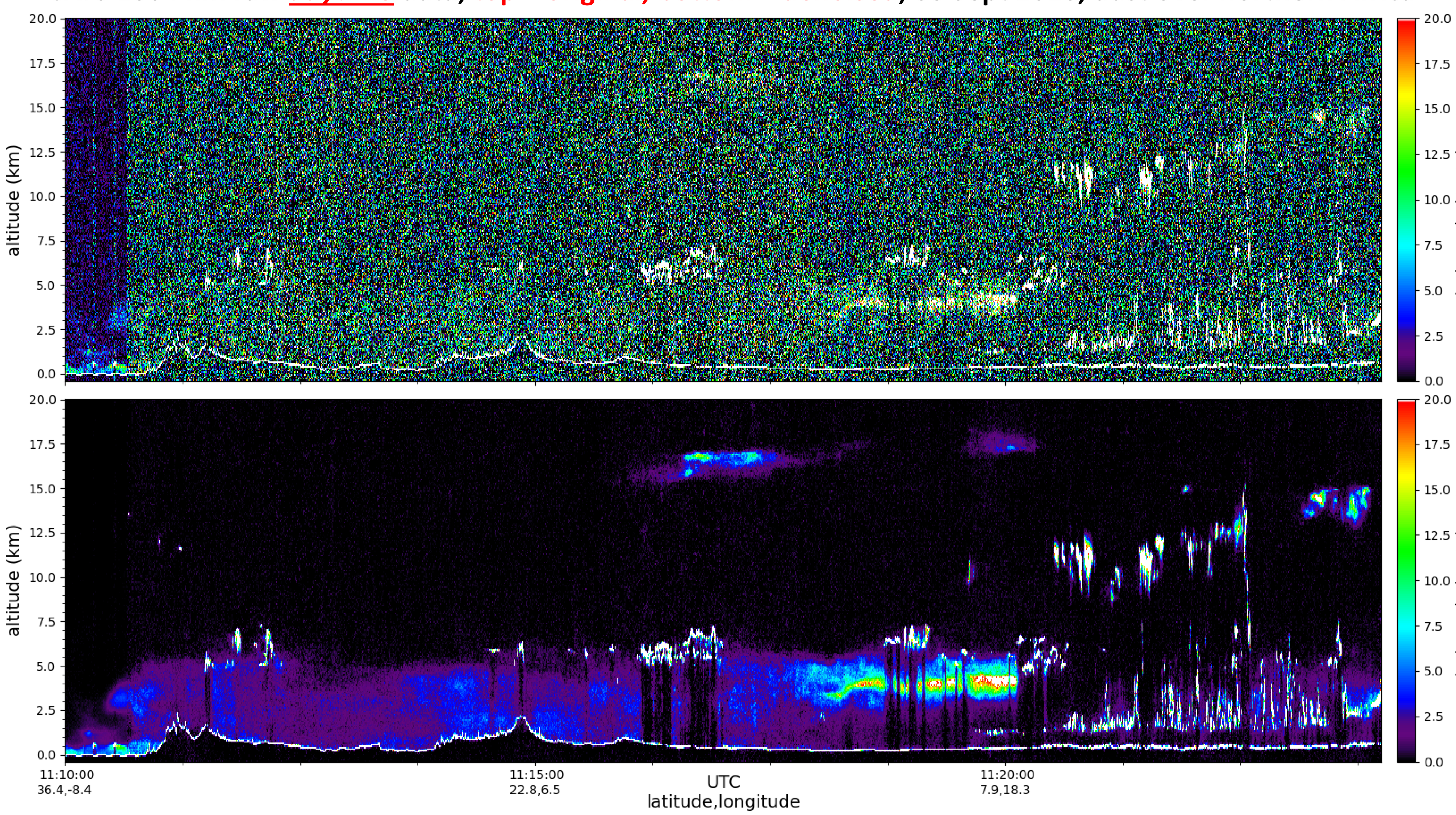
- "DnCNN" [*Zhang et al 2017*]
  - First breakthrough DL image denoising; a series of Conv2D blocks
  - Max depth tried = 17 (Conv2D blocks)
- "DDUNet" [*Jia et al 2021*]
  - More complex learning guidance: cascading UNets, multi-scale processing, and parameter-conservation
  - Max depth tried = 7 (Dense UNet blocks)
  - This architecture yielded best model – ***used in all plots***

TensorFlow

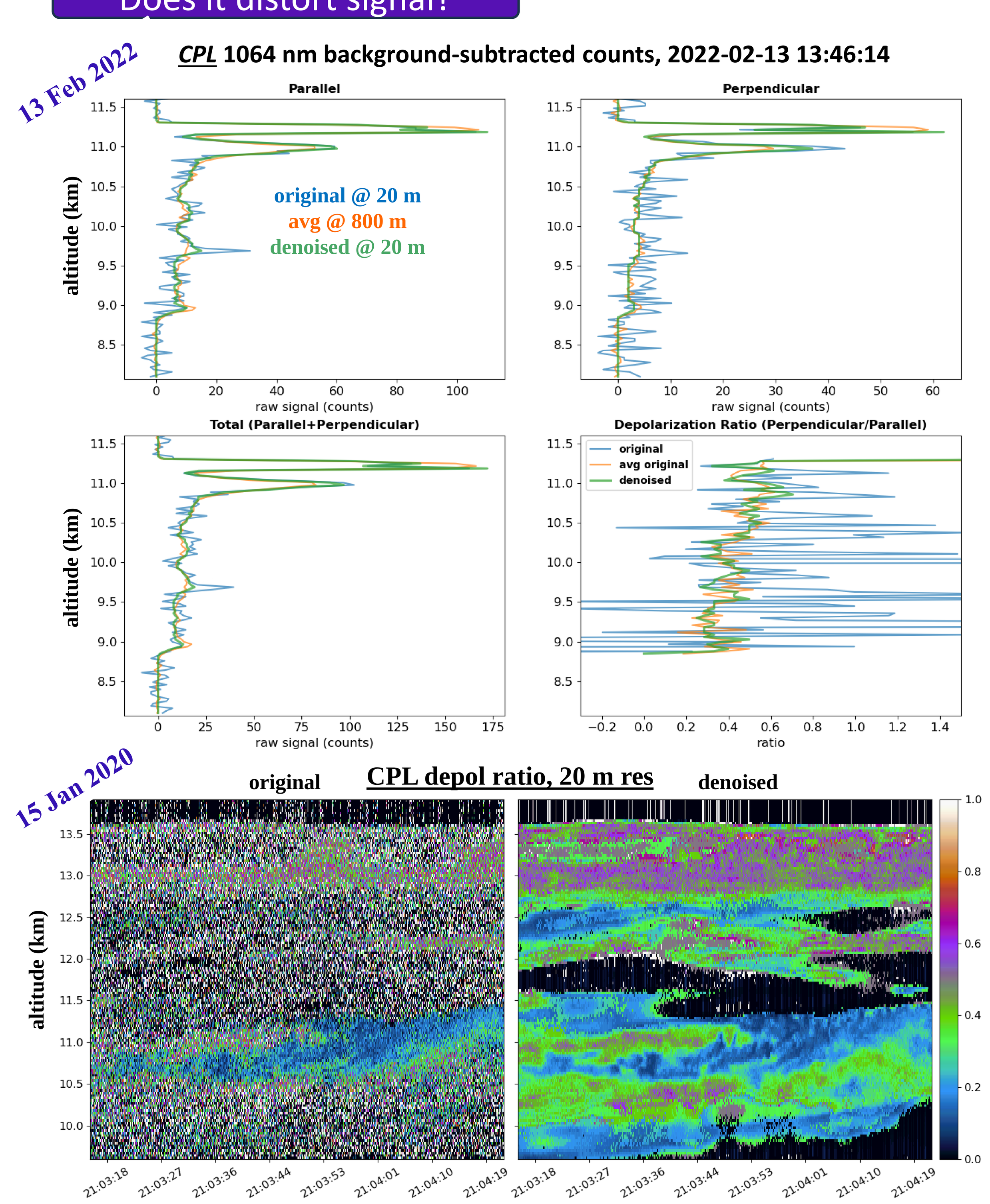
Summary of best model

Data source	CATS raw 1064 nm night data
Training data amount	1.415 months (80%)
Validation data amount	0.354 months (20%)
Poisson noise levels (raw signal counts)	40, 80, & 160
Sample height (bins)	256
Sample width (records)	256
Loss function	L2 loss
Batch size	16
Augmentation used	random flips & translations
Dense UNet blocks	7
Convolutional filters (per block)	64
Trainable parameters	15,528,193
# of GPUs used	4
Training time	~3-4 days
Inference time on 45 minute file	~30 minutes (depends on patch overlap)

CATS 1064 nm raw *daytime* data; top – original, bottom – denoised; 08 Sept 2016, dust over northern Africa

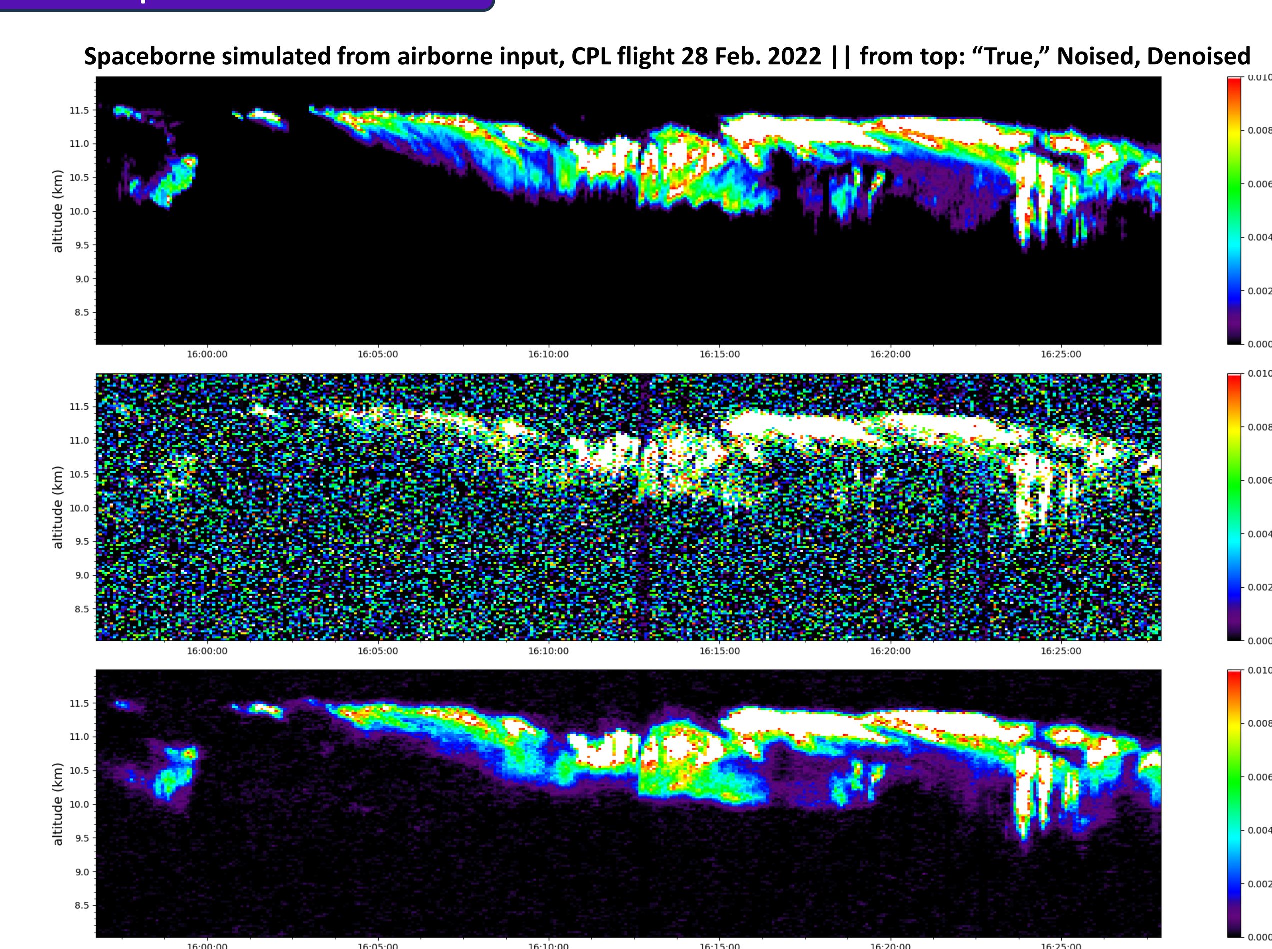


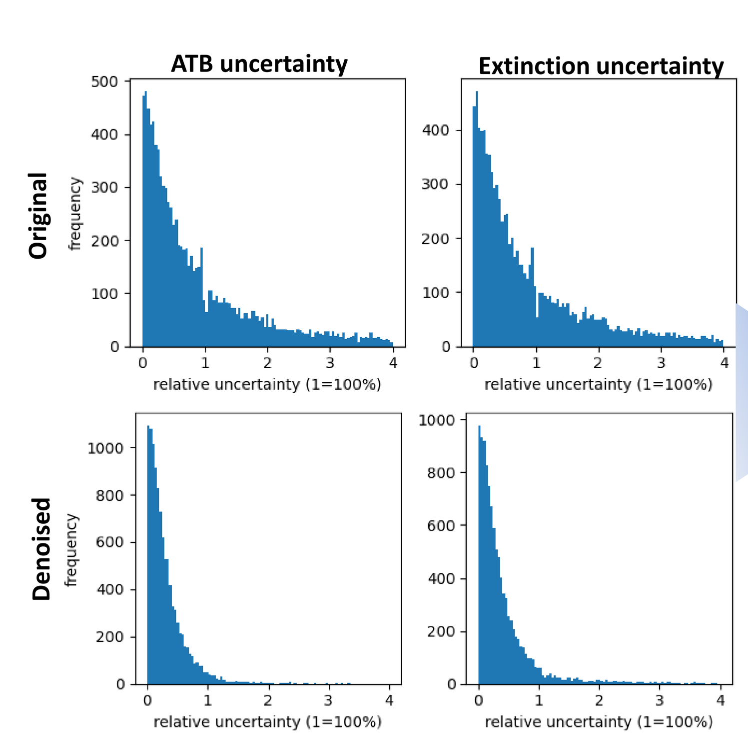
Does it distort signal?



- Model trained exclusively on total signal can effectively denoise both parallel and perpendicular signals, even a different instruments' data
- Visually pleasing fit, no apparent artifacts
- Denoised raw profile corresponds with averaged
- Improved depolarization ratio could allow for more accurate aerosol typing

Does it improve retrievals?





1. Simulated a CATS-like spaceborne lidar signal from high-SNR airborne retrieval
2. Ran standard retrieval algorithm & compared
3. Focused analysis on cirrus cloud depicted above

	Original	*Denoised*
ATB median uncertainty	62%	22%
Ext. median uncertainty	67%	25%
SNR	2.29	4.47

- Limitation: molecular signal rendered undetectable
- Solutions to this limitation are being investigated
- Only random error; assumed perfect calibration and lidar ratio selection

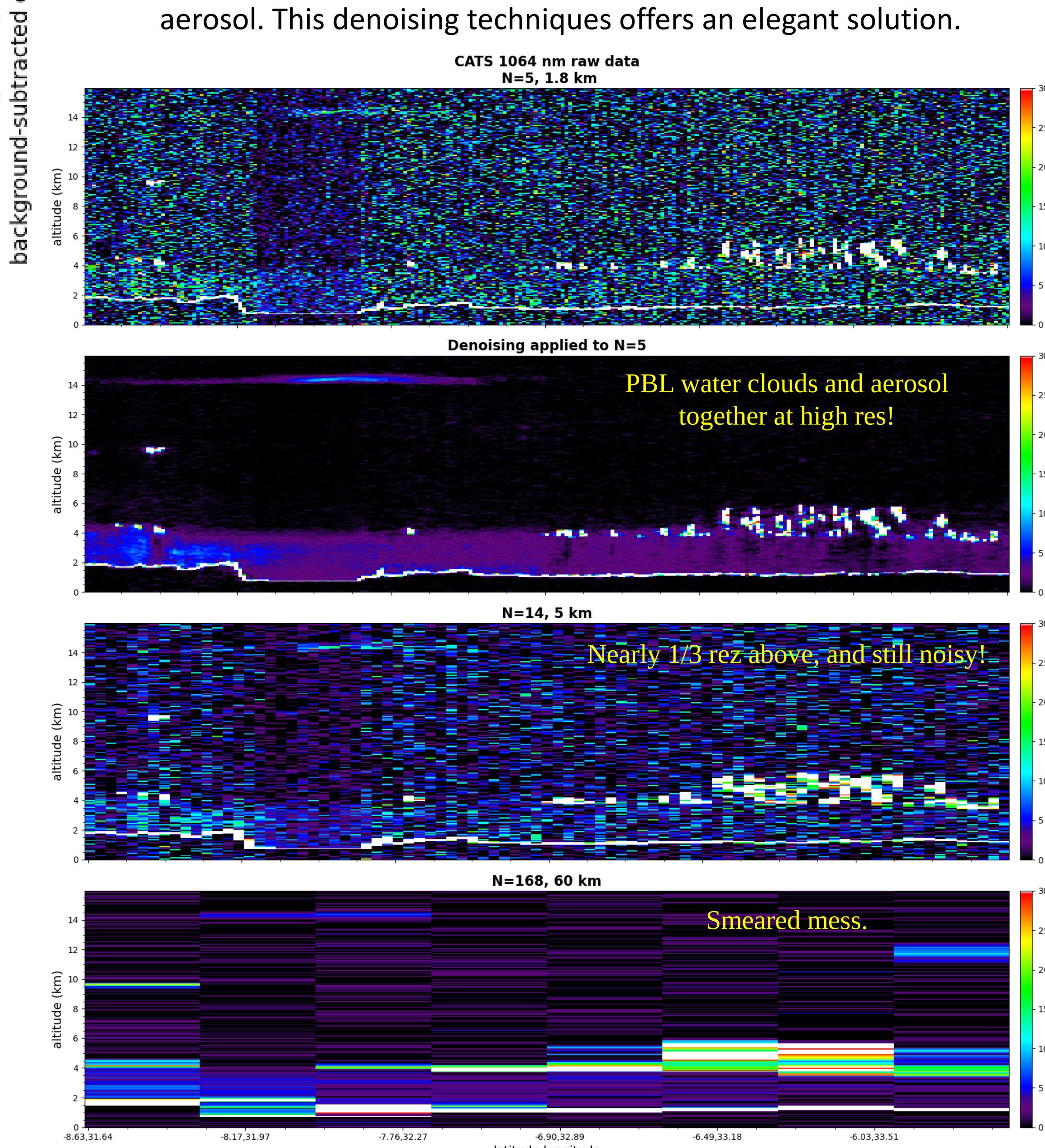
Implications and Future Work

- Better depolarization ratio means better particle typing
- More accurate layer structure at greater resolution
- Better extinction retrievals, better climate science
- Software vs engineering solution
  - Same data quality for less expensive instrument
  - Simply better data with any given instrument
- Applications to several spaceborne lidar datasets

PAST — ICESat-2 ICE, CLOUD, AND LAND ELEVATION SATELLITE-2 ONGOING — NASA EARTH SYSTEM OBSERVATORY AOS ATMOSPHERE OBSERVING SYSTEM FUTURE

No more smearing

CATS Level 2 products averaged to 5 km and 60 km to retrieve aerosol extinction during daytime [*Yorks et al 2020*]. Tiny scale, large signal, low altitude water clouds get smeared with PBL aerosol. This denoising techniques offers an elegant solution.



Next steps

- Work to train an even better model
  - Training set more than double the size has been generated
  - Many training optimization techniques yet to be tested
- More robust performance assessments
- Figure out molecular signal limitation and calibration
- Implement in processing and assess effects on data products

References

Zhang, K.; Zuo, W.; Chen,Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7):3142–3155, 2017.

Jia, F.; Wong, W. H.; Zeng, T. DDUNet: Dense Dense U-Net with Applications in Image Denoising. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 354-364. doi:10.1109/ICCVW54120.2021.00044.

Yorks, J.E.; Selmer, P.A.; Kupchock, A.; Nowottnick, E.P.; Christian, K.E.; Rusinek, D.; Dacic, N.; McGill, M.J. Aerosol and Cloud Detection Using Machine Learning Algorithms and Space-Based Lidar Data. Atmosphere 2021, 12, 606. https://doi.org/10.3390/atmos12050606